THE USE OF THE GENERALIZED INVERSE IN THE GENERAL LINEAR STATISTICAL MODEL

FREDERICK ROBERTS ACKLEY, JR.

THE USE OF THE GENERALIZED INVERSE

IN THE GENERAL LINEAR STATISTICAL MODEL

by

Frederick Roberts Ackley, Jr.
Lieutenant, United States Navy
A.B., Brown University, 1959

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 1967

ABSTRACT

The general linear model of statistical inference is formulated in terms of the Moore-Penrose generalized inverse. The matrix algebra of the generalized inverse which is essential to the model is presented. A methodology for estimation and hypothesis testing is derived which permits identical manipulation of both the full rank and the less-than-full rank cases of the model.

TABLE OF CONTENTS

I. INTRODUCTION

The linear statistical model is a mathematical formulation which is useful in the interpretation of data and observations from experiments. While the assumptions necessary for the use of a linear statistical model restrict its application, it is frequently convenient and appropriate to use such a model because of its simplicity and its suitability to many processes and experiments under investigation.

This paper will be concerned with one of the most frequently utilized linear statistical models. It is assumed that n observations, Y_i, are made of a process or experimental quantity. The process or experiment has p elements, X_j, each of which has a fixed value for each of n replications. Associated with the vector of observations is a vector of errors, denoted by e. The errors are assumed to be uncorrelated between replications and have a multivariate distribution with mean vector 0 and variance-covariance matrix $\sigma^2 I$. In matrix notation, the model is expressed as

$$Y = X\beta + e ,$$

where Y is an n x 1 vector of observations with mean vector $X\beta$ and variance-covariance matrix $\sigma^2 I$. X is an n x p matrix of known constants, and β is a p x 1 vector of unknown parameters. σ^2 is the unknown variance of the individual observations.

By the use of the general linear statistical model, one can estimate the assumed functional relationship between Y and X. If the matrix X is of full rank, we accomplish this by estimation of σ^2 and β. In the less-than-full rank case,

we concern ourselves with estimation of σ^2 and $F'\beta$, sets of linearly independent estimable functions. Several tests of hypotheses are frequently useful in the evaluation of the functional relationships in the model. Such tests provide conclusions concerning the values of the unknown parameters which would be useful in predicting values of Y or in explaining the variability of Y.

This paper discusses the mathematical manipulation of a linear model by which conclusions regarding the process under investigation might be reached. The general linear statistical model has traditionally been treated as two separate cases, with two different methodologies. These separate methodologies are necessitated by two general forms which the matrix X may assume. These forms are the full rank form, where the $n \times p$ matrix X is of rank p, and the less-than-full rank form where X is of rank $r < p$. The application of the general linear model to experimental design frequently gives us the less-than-full rank case, whereas in the regression analysis application, the full rank case is most often encountered.

As an example of the apparent necessity for different methodologies, consider estimation of the vector of unknown parameters. The method of least squares yields the set of equations,

$$X'X\hat{\beta} = X'Y ,$$

where $\hat{\beta}$ is the least squares estimator of the vector β. If X is of full rank, a unique solution for $\hat{\beta}$ exists, because $X'X$ has an inverse. If, however, the rank of X is $r < p$, a

unique solution does not exist, and reparametrization is commonly used to estimate the invariant features of the solution vectors, $\hat{\beta}$. Reparametrization involves the linear transformation of the vector β into the vector α and the consequent change of the model from

$$Y = X\beta + e$$

to

$$Y = Z\alpha + e \ ,$$

where Z is a matrix of full rank. The mathematical manipulation of this model can then be carried out in much the same fashion as the manipulation of the full rank case. The lack of a unified methodology led the author to consider an approach by which both the full rank and the less-than-full rank cases could be manipulated by the same mathematical technique.

In the chapters to follow, manipulations of the general linear statistical model will be formulated in terms of the Moore-Penrose generalized inverse. The definition of the generalized inverse will be taken as the point of departure for Chapter II. A survey of the properties of the generalized inverse which are preliminary to the statistical formulation will be presented.

In Chapter III the estimation of parameters and functions of parameters, and the distribution of the estimates will be discussed. The formulation and testing of hypotheses will be examined in Chapter IV. For a part of the third chapter and the entire fourth chapter it will be necessary to specify a multivariate distribution for the vector of errors. We shall assume that e has the multivariate normal distribution;

that is, $e \sim MVN(0, \sigma^2 I)$. This is a commonly assumed distribution, as it is justifiable for observational errors in a wide range of processes and experiments.

Appendix A is composed of a simple numerical example of the unified methodology derived in the preceding chapters. An experimental design application of the general linear statistical model has been chosen to illustrate the use of the method in the less-than-full rank case.

Throughout the development, it will be noted that strong reliance has been placed upon the notation, terminology, and methods of proof of Graybill (1). This is due, in part, to Graybill's consistency with the generalized inverse formulation. Equally important, however, is the simplicity with which Graybill has presented the basic statistical theory.

II. THE GENERALIZED INVERSE

The mathematical concept known as the generalized inverse of a matrix was first introduced by E. H. Moore in 1920. Moore (2) presented the first published systematic investigation of the properties of the generalized inverse in 1935. The concept was not widely recognized until 1955 when R. A. Penrose (3) independently rediscovered it.

The Moore-Penrose generalized inverse of an arbitrary matrix A has been defined by Penrose (3) as the solution, $G = A^+$ of the following four equations.

$$AGA = A \qquad (2.1)$$

$$GAG = G \qquad (2.2)$$

$$(AG)' = AG \qquad (2.3)$$

$$(GA)' = GA \qquad (2.4)$$

It is convenient to describe the generalized inverse of an arbitrary matrix by two equations which are equivalent to the defining equations. These equations can be formed by substituting eqn. (2.3) into eqn. (2.2), and eqn. (2.4) into eqn. (2.1). Respectively, these relationships are

$$GG'A' = G \qquad (2.5)$$

and

$$AA'G' = A \qquad (2.6)$$

There are several properties of the generalized inverse which will be useful in developing the statistical aspects of this paper. These properties are stated below with source references, when appropriate, where a detailed development of the property may be found.

Property (1). The generalized inverse specified by the four defining equations is unique. [Penrose (3)]

Property (2). $A^{++} = A$. [Penrose (3)]

Property (3). $A^{+'} = A^{'+}$. [Penrose (3)]

Property (4). $A^{+} = A'A^{+'}A^{+}$. [Penrose (3)]

Property (5). $A^{+}A$, AA^{+}, $I-A^{+}A$, and $I-AA^{+}$ are each idempotent. [Penrose (3)]

Property (6). Rank (A^{+}) = rank (A) = rank $(A^{+}A)$ = rank (AA^{+}) = trace $(A^{+}A)$. [Penrose (3)]

Property (7). $(AB)^{+} = B^{+}A^{+}$ if, and only if, $A^{+}A$ and BB' commute, and $A'A$ and BB^{+} commute. [Greville (4)]

Property (8). A necessary and sufficient condition that $AXB = C$ has a solution for X is $AA^{+}CB^{+}B = C$. The general solution is $X = A^{+}CB^{+} + Z - A^{+}AZBB^{+}$, where Z is arbitrary. [Penrose (3)]

Property (9). If A is an $n \times p$ matrix of rank p. $(A'A)^{+} = (A'A)^{-1}$.

Proof: $(A'A)$ has an inverse, for it is a $p \times p$ matrix of rank p. It may easily be verified that $(A'A)^{-1}$ satisfies the defining equations for the generalized inverse of $(A'A)$. As the generalized inverse is unique, $(A'A)^{+} = (A'A)^{-1}$.

Property (10). If A is an $n \times p$ matrix of rank p, $A^{+}A = I$.

Proof: We know from matrix theory that if A is an $n \times p$ matrix of rank p, then A has a left inverse. It can be shown by substitution into the defining equations that the left inverse if A^{+}. The right inverse result for a $p \times n$ matrix of rank p is, similarly, $AA^{+} = I$.

Property (11). If a matrix A is of full rank and is partitioned such that $A = (A_1 \vdots A_2)$ then

$$A^+ = \begin{bmatrix} A_1^+ - A_1^+A_2[(I-A_1A_1^+)A_2]^+ \\ \\ [(I-A_1A_1^+)A_2]^+ \end{bmatrix}. \qquad \text{[Cline (5)]}$$

In particular, it is noted that when $A_1'A_2 = 0$, then

$$A^+ = \begin{bmatrix} A_1^+ \\ \\ A_2^+ \end{bmatrix},$$

since $A_1^+A_2 = (A_1^+A_1^{+\prime}A_1')A_2 = 0.$

III. ESTIMATION

Making the assumptions for the general linear statistical model,

$$Y = X\beta + e,$$

where $E(e) = 0$ and $E(ee') = \sigma^2 I$, provides us with a basis for estimation of the vector of unknown parameters and functions of the parameters. Estimation of these quantities is equivalent to estimation of the relationships which exist between the observations, Y, and the matrix of known constants, X.

As previously indicated, using the method of least squares to minimize $(Y-X\beta)'(Y-X\beta)$ yields the normal equations,

$$X'X\hat{\beta} = X'Y.$$

Regardless of the rank of X, properties (4) and (8) give us a general solution for the least squares estimator of β,

$$\hat{\beta} = (X'X)^+X'Y + [I-(X'X)^+(X'X)]Z = X^+Y + (I-X^+X)Z,$$

where Z is an arbitrary p x 1 vector. It is apparent that if X is of full rank, then by properties (9) and (10),

$$\hat{\beta} = X^+Y = (X'X)^{-1}X'Y,$$

which confirms the result derived in Chapter I. However, we will not set the full rank case apart, for the general solution holds for rank $(X) = r \leq p$.

Least squares estimation does not directly provide an estimate of σ^2. However, by basing our estimate of σ^2 on $\hat{\beta}$, an unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{n - r} = \frac{Y'(I-XX^+)Y}{n - r}$$

Of immediate interest is the mean vector and variance-covariance matrix of the estimator $\hat{\beta}$. Recalling that Y has

mean vector $X\beta$ and variance-covariance matrix $\sigma^2 I$,

$$E(\hat{\beta}) = E[X^+Y+(I-X^+X)Z] = X^+X\beta + [I-X^+X]Z$$

and

$$\mathrm{Var}(\hat{\beta}) = E[(X^+Y-X^+X\beta)(X^+Y-X^+X\beta)'] = E[X^+ee'X^{+\prime}] = \sigma^2(X'X)^+.$$

It is noted that $\hat{\beta}$ is not, in general, an unbiased estimator. However, if the vector β happens to equal our choice of the arbitrary vector Z, the estimator $\hat{\beta}$ is unbiased.

If we now specify the distribution of the vector of errors such that $e \sim MVN(0,\sigma^2 I)$, it can be shown that the method of maximum likelihood estimation leads to precisely the same results as the least squares estimation of β.

By assuming that $E(e) = 0$ and $E(ee') = \sigma^2 V$, where V is an arbitrary positive definite matrix, the estimation can be somewhat generalized. The normal equations stemming from least squares estimation in this more general case are,

$$(X'V^{-1}X)\hat{\beta} = X'V^{-1}Y,$$

with solutions,

$$\hat{\beta} = (X'V^{-1}X)^+X'V^{-1}Y + [I-(X'V^{-1}X)^+(X'V^{-1}X)]Z.$$

Furthermore,

$$E(\hat{\beta}) = (X'V^{-1}X)^+(X'V^{-1}X)\beta + [I-(X'V^{-1}X)^+(X'V^{-1}X)]Z$$

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= E[(\hat{\beta}-E(\hat{\beta}))(\hat{\beta}-E(\hat{\beta}))'] \\ &= E[(X'V^{-1}X)^+X'V^{-1}ee'V^{-1}X(X'V^{-1}X)^+] \\ &= (X'V^{-1}X)^+(X'V^{-1}X)(X'V^{-1}X)^+\sigma^2 = \sigma^2(X'V^{-1}X)^+.\end{aligned}$$

In each of the distributional cases discussed above, the estimator of β is not unique if the rank $(X) = r < p$. In the less-than-full rank case no direct inference can be made concerning all the values of the elements of β. However, there

are invariant quantities in any solution which are called estimable functions.

Consider a linear combination of the elements of $\hat{\beta}$, denoted by $f'\hat{\beta}$. $f'\hat{\beta}$ is invariant with respect to the arbitrary Z if, and only if,

$$f'[I-X^+X] = 0.$$

If X is of full rank, then this relationship holds for all f'. In the less-than-full rank case, all linear combinations of the form where $f' = b'X$ have this property since

$$f'[I-X^+X] = b'[X-XX^+X] = 0.$$

Furthermore, since the nullity of $[I-X^+X]$ is r, any basis for the null space of $[I-X^+X]$ is a set of r linearly independent solutions to $f'[I-X^+X] = 0$. Since the rank of X is r and $b'X[I-X^+X] = 0$, all such vectors f' can be written $f' = b'X$. We may define a linearly estimable function in terms of this invariance property; that is, a linear function of the parameters, $f'\beta$, is estimable if, and only if, $f'(I-X^+X) = 0$. Graybill (1) defines a linearly estimable function as a function of the unknown parameters for which there exists a vector b' such that

$$E(b'Y) = b'X\beta = f'\beta.$$

Clearly, the two definitions are equivalent. Graybill furthermore proves that it is equivalent to say that a linearly estimable function is a linear combination of the parameters for which a solution for r exists to the equation

$$f' = r'X'X.$$

It is this form of the estimable function which will be most useful to us.

Among other results which Graybill has developed with applicability in the generalized inverse formulation are the three listed below:

(1). The best linear unbiased estimator of an estimable function, $f'\beta$, is

$$\widehat{f'\beta} = f'\hat{\beta}$$

(2). The functions, $f_1'\beta$, $f_2'\beta$, ..., $f_m'\beta$ are called linearly independent estimable functions if each of the functions is linearly estimable, and f_1', f_2', ..., f_m' are linearly independent vectors.

(3). A set of linearly independent estimable functions contains, at most, r functions, where $r = \text{rank}\ (X)$.

A linearly estimable matrix function, $F'\beta$, is a set of m linearly independent estimable functions, $1 \leq m \leq r$. A matrix function may be represented as

$$F'\beta = \begin{bmatrix} f_1' \\ f_2' \\ \cdot \\ \cdot \\ \cdot \\ f_m' \end{bmatrix} (\beta) = (R'X'X)\beta$$

where the $f_i'\beta$ are linearly independent estimable functions.

A result which will be frequently utilized is that the estimator of an estimable matrix function is invariant with respect to the arbitrary matrix Z, which follows directly from the fact that each component of $F'\hat{\beta}$ has this property.

Considering the distributional case where $e \sim MVN(0,\sigma^2 I)$, the estimator of an estimable matrix function is always unique and, therefore, of use in making inferences regarding the values of the estimable functions. The mean vector and variance-covariance matrix of our estimator are, respectively,

$$E(F'\hat{\beta}) = E(F'X^+Y) = R'X'(XX^+X)\beta = F'\beta$$

and

$$Var(F'\hat{\beta}) = E[(F'X^+Y-F'\beta)(F'X^+Y-F'\beta)'] = E(F'X^+ee'X^{+\prime}F) = \sigma^2 F'F.$$

IV. TESTS OF HYPOTHESES

We shall now discuss a test of the hypothesis $F'\beta = C$, where $F'\beta$ is an m x 1 vector of linearly independent estimable functions. $F'\beta = C$ is called a linearly estimable hypothesis if $F'\beta$ is a set of linearly independent estimable functions and C is a vector of known constants. To illustrate this concept, suppose we desire to test the equality of the first three components of β; that is,

$$H_0:\beta_1 = \beta_2 = \beta_3.$$

If the functions,

$$f_1'\beta = (1 \ -1 \ 0 \ 0\cdots 0)\beta$$

and

$$f_2'\beta = (1 \ 1 \ -2 \ 0\cdots 0)\beta$$

are linearly estimable, then

$$F'\beta = (f_1', f_2')'\beta = \begin{bmatrix} \beta_1 - \beta_2 \\ \\ \beta_1 + \beta_2 - 2\beta_3 \end{bmatrix} = 0$$

if, and only if, $\beta_1 = \beta_2 = \beta_3$. We require that the linear estimable functions of the hypothesis be independent. To have any of these not independent would be redundant. For example, in the illustration above,

$$f_3'\beta = (0 \ 1 \ -1 \ 0\cdots 0)\beta = 0$$

would test the equality of β_2 and β_3, which is already being tested by $f_1'\beta$ and $f_2'\beta$.

In the full rank case, a consequence of the form of the linearly estimable function and the estimable hypothesis is that the elements of β, individually, and as a complete set, are estimable. The hypothesis that $\beta = C$ is then estimable.

Furthermore, any subset of the elements of β equal to a specified set of constants may be tested by choosing the appropriate rows of the p x p identity matrix as the matrix F'.

In developing the likelihood ratio test criterion, we will be concerned only with the normal theory case where $e \sim MVN(0, \sigma^2 I)$. The appropriate likelihood function is then

$$f(e;\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{-(Y-X\beta)'(Y-X\beta)}{2\sigma^2}\right].$$

To develop a test of the hypothesis H_0: $F'\beta = C$ versus the alternative, H_1:$F'\beta \neq C$, we will utilize the likelihood ratio,

$$\theta = \frac{L(\hat{\omega})}{L(\hat{\Omega})}.$$

$L(\hat{\Omega})$ is the maximum value of the likelihood function with the parameters contained in the (β,σ^2) p + 1 dimensional space, Ω, which is unrestricted. $L(\hat{\omega})$ is similarly the maximum value of the likelihood function with the parameter space restricted by H_0. The ω space is p - m + 1 dimensional because values for m independent relationships among the elements of β are specified by the hypothesis. The maximizing values of the parameters in the ω space shall be denoted by $\tilde{\beta}$ and $\tilde{\sigma}^2$, while the corresponding values in the Ω space shall be denoted by $\hat{\beta}$ and $\hat{\sigma}^2$.

In the restricted parameter space we desire to maximize $f(e;\beta,\sigma^2)$ with respect to σ^2 and β subject to the restraint, $F'\beta = C$.

Let

$$\phi(e;\beta,\sigma^2) = \log_e[f(e;\beta,\sigma^2)].$$

It follows that

$$\frac{\partial\phi}{\partial\sigma^2} = \frac{(Y-X\beta)'(Y-X\beta)}{2\tilde{\sigma}^4} - \frac{n}{2\tilde{\sigma}^2} = 0.$$

The resulting maximizing value of σ^2 for any given β is

$$\tilde{\sigma}^2 = \frac{(Y-X\beta)'(Y-X\beta)}{n} .$$

Examining the likelihood function it is noted that the maximum value of the likelihood function with respect to β with $\sigma^2 = \tilde{\sigma}^2 = (Y-X\beta)'(Y-X\beta)/n$ occurs when $(Y-X\beta)'(Y-X\beta)$ is a minimum. However, we must constrain β by the relationship, $F'\beta = C$. The problem of finding the value of β which maximizes the likelihood function within the hypothesis constraint can then be formulated as the quadratic program,

$$\min\ (Y-X\beta)'(Y-X\beta),$$

$$\text{subject to: } F'\beta = C.$$

The derivatives of the Lagrangian of the objective function are:

$$\frac{\partial\Psi}{\partial\beta} = -2Y'X + 2\beta'X'X - 2\lambda'F' = 0 \qquad (4.1)$$

$$\frac{\partial\Psi}{\partial\lambda} = F'\beta - C = 0, \qquad (4.2)$$

which are the necessary and sufficient conditions for a minimization where $2\lambda'$ is the appropriate vector of Lagrange multipliers. We first premultiply eqn. (4.1) by $AX^{+}{}'$, yielding

$$-AY + F'\beta - AA'\lambda = 0.$$

Substituting C, from eqn. (4.2), for $F'\beta$, the solution for λ is

$$\lambda = (AA')^{+}C - A^{+}{}'Y.$$

Returning to eqn. (4.1) with this value of λ, and solving for $\tilde{\beta}$, yields

$$\tilde{\beta} = (X'X)^{+}X'A'(A'^{+}A^{+}C-A'^{+}Y) + X^{+}X'^{+}X'Y + (I-X^{+}X)Q$$
$$= X^{+}A^{+}C-X^{+}A^{+}AY + X^{+}Y + (I-X^{+}X)Q, \qquad (4.3)$$

where Q is an arbitrary p x 1 vector.

Since

$$-2X'Y+2X'X[X^{+}A^{+}C-X^{+}A^{+}AY+X^{+}Y+(I-X^{+}X)Q]-2X'A'[(AA')^{+}C-A^{+}{}'Y] = 0,$$

$$-X'Y+X'A^{+}C-X'A^{+}AY+X'Y-X'A^{+}C+X'A'A^{+}{}'Y = 0,$$

clearly, eqn. (4.1) is satisfied by $\tilde{\beta}$ and λ. Also

$$F'\beta = R'X'X[X^{+}A^{+}C-X^{+}A^{+}AY+X^{+}Y+(I-X^{+}X)Q]$$
$$= AA^{+}C-AA^{+}AY+AY.$$

Since A is an m x n matrix of rank m,

$$AA^{+} = I,$$

and eqn. (4.2) is also satisfied.

Returning to the likelihood function and substituting $\tilde{\sigma}^2$, while retaining the general term, $\tilde{\beta}$, we find that

$$L(\hat{\omega}) = \max_{\substack{\sigma^2 \varepsilon \omega \\ F'\beta = C}} f(e;\beta,\sigma^2) = \frac{n^{n/2}\exp[-n/2]}{(2\pi)^{n/2}[(Y-X\tilde{\beta})'(Y-X\tilde{\beta})]^{n/2}}$$

Finding the maximizing values of σ^2 and β in the unrestricted parameter space yields the results,

$$\hat{\beta} = X^{+}Y + (I-X^{+}X)Z \qquad (4.4)$$
$$\hat{\sigma}^2 = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{n},$$

where Z is an arbitrary p x 1 vector, and $\hat{\sigma}^2$ is a biased estimator of σ^2. Combining these results in the likelihood function yields the maximum value,

$$L(\hat{\Omega}) = \max_{\substack{\sigma^2 \varepsilon \Omega \\ \beta \varepsilon \Omega}} f(e;\beta,\sigma^2) = \frac{n^{n/2}\exp[-n/2]}{(2\pi)^{n/2}[(Y-X\hat{\beta})'(Y-X\hat{\beta})]^{n/2}} .$$

The likelihood ratio is then

$$\theta = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{(Y-X\tilde{\beta})'(Y-X\tilde{\beta})}\right]^{n/2} ,$$

where $\tilde{\beta}$ and $\hat{\beta}$ are as presented in equations (4.3) and (4.4) respectively.

To develop our test criterion we will first manipulate the quadratic forms of the likelihood ratio to form chi-square random variables. The independence of the quadratic forms will then be demonstrated and the test criterion established as a non-central F statistic.

For any $\tilde{\beta}$,

$$(Y-X\tilde{\beta})'(Y-X\tilde{\beta}) = (Y-X\hat{\beta})'(Y-X\hat{\beta}) + (\hat{\beta}-\tilde{\beta})'(X'X)(\hat{\beta}-\tilde{\beta}).$$

As a consequence, the likelihood ratio can be written as

$$\theta = \left[\frac{1}{1+\frac{(\hat{\beta}-\tilde{\beta})'X'X(\hat{\beta}-\tilde{\beta})}{(Y-X\hat{\beta})'(Y-X\hat{\beta})}}\right]^{n/2} . \qquad (4.5)$$

Let us now examine the ratio in the denominator. Using the solutions for $\hat{\beta}$ and $\tilde{\beta}$, it follows that

$$\begin{aligned}
&(\hat{\beta}-\tilde{\beta})X'X(\hat{\beta}-\tilde{\beta}) \\
&\quad = (XX^+Y-XX^+A^+C+XX^+A^+AY-XX^+Y)'(XX^+Y-XX^+A^+C+XX^+A^+AY-XX^+Y) \\
&\quad = (XX^+A^+AY-XX^+A^+C)'(XX^+A^+AY-XX^+A^+C) \\
&\quad = (AY-C)'(A^{+\prime}XX^+A^+)(AY-C).
\end{aligned}$$

By the definition of a non-central chi-square ($\chi^{\prime 2}$) random variable, if $Z \sim MVN(\mu,V)$, then $Z'BZ \sim \chi^{\prime 2}(q,\lambda)$, where q is the rank of BV and $\lambda = \mu'B\mu/2$, if, and only if, BV is

idempotent. Then if $[(A^{+\prime}XX^{+}A^{+})/\sigma^{2}][\mathrm{Var}(AY-C)]$ is idempotent, $(\hat{\beta}-\tilde{\beta})'X'X(\hat{\beta}-\tilde{\beta})/\sigma^{2}$ has a non-central chi-square distribution. Since $Y \sim MVN(X\beta,\sigma^{2}I)$, $(AY-C) \sim MVN(AX\beta-C,\sigma^{2}AA')$ and

$$BV = \frac{A^{+\prime}XX^{+}A^{+}}{\sigma^{2}}\,\sigma^{2}AA'.$$

We must determine whether, or not, $A^{+\prime}XX^{+}A^{+}AA'$ is idempotent. It is clear that

$$A^{+\prime}XX^{+}(A^{+}AA') = A^{+\prime}XX^{+}(A') = A^{+\prime}XX^{+}(XR),$$

where $R'X'X = AX$. Therefore,

$$A^{+\prime}XX^{+}A^{+}AA' = A^{+\prime}XR = A^{+\prime}A'.$$

Since A' is a $p \times m$ matrix of rank m, it has a left inverse, $A^{+\prime}$, and

$$BV = A^{+\prime}XX^{+}A^{+}AA' = I.$$

Hence, BV is idempotent and of rank m. We can therefore conclude that $(AY-C)'(A^{+\prime}XX^{+}A^{+})(AY-C)/\sigma^{2}$ has the non-central chi-square distribution with m degrees of freedom and non-centrality parameter, $\lambda = (F'\beta-C)'(A^{+\prime}XX^{+}A^{+})(F'\beta-C)/2\sigma^{2}$.

The quadratic form, $(Y-X\hat{\beta})'(Y-X\hat{\beta})$ is equal to $Y'(I-XX^{+})Y$. Since $(I-XX^{+})$ is idempotent of rank $n-r$, $(Y-X\hat{\beta})'(Y-X\hat{\beta})$ has the central chi-square distribution with $n-r$ degrees of freedom. Consequently, our likelihood ratio is

$$\theta = \left[\frac{1}{1+\frac{\gamma}{\zeta}}\right]^{n/2} \tag{4.6}$$

where $\gamma \sim \chi'^{2}(m,(F'\beta-C)'(A^{+\prime}XX^{+}A^{+})(F'\beta-C)/2\sigma^{2})$ and the variable $\zeta \sim \chi^{2}(n-r)$.

In general, two quadratic forms $T'DT$ and $T'ET$ are independent if, and only if, $DVE = 0$, where $T \sim MVN(\mu,V)$.

The identities,

$$\gamma = (\hat{\beta}-\tilde{\beta})'X'\chi(\hat{\beta}-\tilde{\beta})/\sigma^2 = (Y-X\tilde{\beta})XX^+(Y-X\tilde{\beta})/\sigma^2,$$

and

$$\zeta = (Y-X\hat{\beta})'(Y-X\hat{\beta})/\sigma^2 = (Y-X\tilde{\beta})'(I-XX^+)(Y-X\tilde{\beta})/\sigma^2$$

are convenient forms for demonstrating the independence of γ and ζ. Since

$$(Y-X\tilde{\beta}) = (I-XX^++XX^+A^+A)Y-XX^+A^+C\ ,$$

and the variance of Y is $\sigma^2 I$, then

$$\mathrm{Var}(Y-X\tilde{\beta}) = \sigma^2(I-XX^++XX^+A^+A)(I-XX^++XX^+A^+A)' = \sigma^2(I-XX^++A^+A).$$

The quantity corresponding to DVE is then

$$\left(\frac{X'X}{\sigma^2}\right)(I-XX^++A^+A)\left(\frac{I-XX^+}{\sigma^2}\right) = 0,$$

as can be easily verified.

The ratio of the non-central chi-square variable to the central chi-square variable in the denominator of our likelihood ratio with each variable divided by its degrees of freedom has the non-central F distribution with m and n-r degrees of freedom and non-centrality parameter,

$$\lambda = (F'\beta-C)'(A^{+\prime}XX^+A^+)(F'\beta-C).$$

A necessary and sufficient condition that $\lambda = 0$ is that the hypothesis is true; that is, $F'\beta-C = 0$. In this event, the ratio of the quadratic forms divided by their degrees of freedom has a central F distribution,

$$\frac{\gamma(n-r)}{\zeta\ (m)} = \frac{(AY-C)'(A'^+XX^+A^+)(AY-C)}{(Y-X\hat{\beta})'(Y-X\hat{\beta})} \sim F_{m,n-r}\ .$$

Since rejection of the hypothesis is consistent with small values of the likelihood ratio and the likelihood ratio is monotonic decreasing in $[\gamma(n-r)]/[\zeta(m))]$, the critical region

for this test is

$$\frac{\gamma(n-r)}{\zeta\ (m)} > F_{m,n-r}(\alpha),$$

where $F_{m,n-r}\ (\alpha)$ is the upper significance point corresponding to significance level α.

As substantiation of our likelihood ratio criterion, we shall compare it with commonly used criteria for two types of tests in the full rank case. The common form of the full hypothesis test criterion for $\beta = \beta^*$, Graybill (1) for example, is

$$\frac{n-p}{p}\ \frac{(Y-X\beta^*)(X(X'X)^{-1}X')(Y-X\beta^*)}{(Y-XX^+Y)'(Y-XX^+Y)} > F_{p,n-p}(\alpha) \quad .$$

It is apparent that the denominators are equal. Furthermore,

$$\begin{aligned}(Y-X\beta^*)'(X(X'X)^{-1}X')(Y-X\beta^*) &= (Y-X\beta^*)'(XX^+)'(XX^+)(Y-X\beta^*) \\ &= (X^+Y-\beta^*)'(X'X)(X^+Y-\beta^*) \quad .\end{aligned}$$

Therefore the test criteria are equivalent in the full hypothesis test.

To test that s of the elements of β are equal to known constants, while the other p - s elements are unspecified, Graybill (1) uses the subhypothesis criterion,

$$\frac{n-p}{s}\ \frac{(Y-X_1\gamma_1^*)'(X(X'X)^{-1}X'-X_2(X_2'X_2)^{-1}X_2')(Y-X_1\gamma_1^*)}{(Y-X_1\gamma_1^*)'(I-X(X'X)^{-1}X')(Y-X_1\gamma_1^*)} > F_{s,n-p}(\alpha).$$

For the purpose of this full rank test, the elements of the general linear model have been partitioned such that

$$X = (X_1,X_2),\ \beta' = (\gamma_1,\gamma_2)$$

and the hypothesis is $\gamma_1 = \gamma_1^*$, where γ_1^* is an s x 1 vector of known constants.

In the generalized inverse formulation this hypothesis may be tested by setting $C = \gamma_1^*$ and choosing F' of the form $F' = [I \vdots 0]$, where the identity matrix is of dimension s and the null matrix has dimensions $s \times (p-s)$. As a simplification we shall consider only the full rank case where γ_1 is orthogonal to γ_2; that is, $X_1'X_2 = 0$. Since $F' = AX = [I \vdots 0]$, we must have

$$AX_1 = I$$

$$AX_2 = 0.$$

By property (11) of Chapter II, it is then clear that $A = X_1^+$. Our hypothesis is then $F'\beta = \gamma_1^*$, and the test criterion is

$$\frac{n-p}{s} \quad \frac{(X_1^+Y-\gamma_1^*)'(X_1'XX^+X_1)(X_1^+Y-\gamma_1^*)}{(Y-X\hat{\beta})'(Y-X\hat{\beta})} > F_{s,n-p}(\alpha) \ .$$

Again it is apparent that the denominators are equal. The numerators are also equal since

$$\begin{aligned}
&(Y-X_1\gamma_1^*)'(X(X'X)^{-1}X'-X_2(X_2'X_2)^{-1}X_2')(Y-X_1\gamma_1^*) \\
&= (Y-X_1\gamma_1^*)'(XX^+-X_2X_2^+)(Y-X_1\gamma_1^*) \\
&= (Y-X_1\gamma_1^*)'(X_1X_1^+)(Y-X_1\gamma_1^*) \\
&= (Y-X_1\gamma_1^*)'X_1'^+(X_1'XX^+X_1)X_1^+(Y-X_1\gamma_1^*) \\
&= (X_1^+Y-\gamma_1^*)'(X_1'XX^+X_1)(X_1^+Y-\gamma_1^*) \ .
\end{aligned}$$

The equivalence of our test criterion to the commonly used criterion in this special case of the subhypothesis test is then clear.

BIBLIOGRAPHY

1. Graybill, Franklin A. An Introduction to Linear Statistical Models. Volume I. New York: McGraw-Hill Book Company, Inc., 1961.

2. Moore, Eliakim M. General Analysis. Philadelphia: The American Philosophical Society, 1935.

3. Penrose, R. "A Generalized Inverse for Matrices," Proceedings of the Cambridge Philosophical Society 51:406-13, 1955.

4. Greville, T.N.E. "Note on the Generalized Inverse of a Matrix Product," SIAM Review, 8:518-21, 1966.

5. Cline, Randall E. "Representations for the Generalized Inverse of a Partitioned Matrix," Journal SIAM, 12:588-600, 1964.

APPENDIX A

Estimation and Hypothesis Testing - An Illustration

Consider a completely randomized experimental design model, where

$$Y_{ij} = \mu + \tau_i + e_{ij} \qquad i = 1,2,3$$
$$j = 1,2$$

We shall suppose that the required normality assumptions are justifiable for our statistical model,

$$Y = X\beta + e$$

where

$$e \sim MVN(0,\sigma^2 I).$$

The necessary elements of our statistical model are:

Observations — Design Structure

$$Y = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 2 \\ 8 \\ 10 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

The rank of X is $r = 3$. By solution of the equations defining the generalized inverse,

$$X^+ = 1/8 \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & -1 & -1 & -1 & -1 \\ -1 & -1 & 3 & 3 & -1 & -1 \\ -1 & -1 & -1 & -1 & 3 & 3 \end{vmatrix}$$

$$\text{and} \quad (X'X)^{+} = 1/64 \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 22 & -10 & -10 \\ 2 & -10 & 22 & -10 \\ 2 & -10 & -10 & 22 \end{bmatrix}$$

As this is an example of the less-than-full rank case, the estimate is not unique. The estimator has the general solution,

$$\hat{\beta} = X^{+}Y + [I-X^{+}X]Z$$

$$= 1/8 \begin{bmatrix} 26 \\ -10 \\ -10 \\ 46 \end{bmatrix} + 1/8 \begin{bmatrix} 2 & -2 & -2 & -2 \\ -2 & 2 & 2 & 2 \\ -2 & 2 & 2 & 2 \\ -2 & 2 & 2 & 2 \end{bmatrix} Z$$

The mean vector of $\hat{\beta}$ is

$$E(\hat{\beta}) = X^{+}X\beta = [I-X^{+}X]Z$$

$$= 1/8 \left[\begin{bmatrix} 6\mu+2\tau_1+2\tau_2+2\tau_3 \\ 2\mu+6\tau_1-2\tau_2-2\tau_3 \\ 2\mu-2\tau_1+6\tau_2-2\tau_3 \\ 2\mu-2\tau_1-2\tau_2+6\tau_3 \end{bmatrix} + \begin{bmatrix} 2Z_1-2Z_2-2Z_3-2Z_4 \\ -2Z_1+2Z_2+2Z_3+2Z_4 \\ -2Z_1+2Z_2+2Z_3+2Z_4 \\ -2Z_1+2Z_2+2Z_3+2Z_4 \end{bmatrix} \right]$$

and the variance-covariance matrix of $\hat{\beta}$ is

$$\text{Var}\ (\hat{\beta}) = \sigma^2 (X'X)^{+} = \sigma^2/32 \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 11 & -5 & -5 \\ 1 & -5 & 11 & -5 \\ 1 & -5 & -5 & 11 \end{bmatrix}$$

Furthermore, the unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{n-r} = 4/3$$

The estimator of β in the less-than-full rank case by itself does not tell us anything about the individual components of β. However, consider the estimate of a function of parameters, where

$$R' = \begin{pmatrix} 0 & 1/2 & -1/2 & 0 \\ 0 & 1/2 & 1/2 & -1 \end{pmatrix},$$

$$A = R'X' = \begin{pmatrix} 1/2 & 1/2 & -1/2 & -1/2 & 0 & 0 \\ 1/2 & 1/2 & 1/2 & 1/2 & -1 & -1 \end{pmatrix},$$

and

$$F' = AX = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{pmatrix}.$$

The estimator of the estimable matrix function, $F'\beta$, is

$$F'\hat{\beta} = FX^{+}Y = \begin{pmatrix} 0 \\ 14 \end{pmatrix}$$

with mean and variance-covariance matrix

$$E(F'\hat{\beta}) = F'\beta = \begin{pmatrix} \tau_1 - \tau_2 \\ \tau_1 + \tau_2 - 2\tau_3 \end{pmatrix}$$

$$\mathrm{Var}\,(F'\beta) = \sigma^2 F'F = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}.$$

It is commonly desired to test the equality of all the τ_i, the treatment effects. In this event, the corresponding estimable hypothesis is

$$F'\beta = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{pmatrix} \beta = \begin{pmatrix} \tau_1 - \tau_2 \\ \tau_1 + \tau_2 - 2\tau_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

For this estimable hypothesis, the generalized inverse of A' is

$$A^{+\prime} = \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & -1/3 & -1/3 \end{bmatrix}.$$

The test statistic,

$$\left(\frac{n-r}{m}\right) \frac{\gamma}{\zeta} = \left(\frac{n-r}{m}\right) \frac{(AY-C)'(A^{+\prime}XX^{+}A^{+})(AY-C)}{(Y-X\hat{\beta})'(Y-X\hat{\beta})} = 24.5.$$

The critical region for this test is

$$\left(\frac{n-r}{m}\right) \frac{\gamma}{\zeta} > F_{2,3}(.05) = 9.55.$$

We must therefore reject the hypothesis, H_0: $\tau_1 = \tau_2 = \tau_3$.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library Naval Postgraduate School Monterey, California 93940	2
3.	Chief of Naval Operations (OP 96) Navy Department Washington, D.C. 20350	1
4.	Prof. Harold J. Larson Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	5
5.	Prof. Donald R. Barr Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
6.	Prof. Jack R. Borsting Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
7.	Lt. Frederick R. Ackley, Jr., USN 431 Raymond Street Rockville Centre, N. Y. 11570	1

Unclassified
Security Classification

UNCLASSIFIED

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

The Use of the Generalized Inverse in the General Linear Statistical Model

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Thesis

5. AUTHOR(S) *(Last name, first name, initial)*

Ackley, Frederick Roberts, Jr., Lieutenant, USN

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1967	32	5

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. AVAILABILITY/LIMITATION NOTICES

[redacted]

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California

13. ABSTRACT

The general linear model of statistical inference is formulated in terms of the Moore-Penrose generalized inverse. The matrix algebra of the generalized inverse which is essential to the model is presented. A methodology for estimation and hypothesis testing is derived which permits identical manipulation of both the full rank and the less-than-full rank cases of the model.

UNCLASSIFIED

DD FORM 1 JAN 64 1473

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
General linear statistical model						
Generalized inverse						
Statistical estimation						
Hypothesis testing						

UNCLASSIFIED

[redacted]

[redacted]

UNCLASSIFIED